Geoderma 314 (2018) 146–159

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach

Anukool Raj[a], Somsubhra Chakraborty[a], Bogdan M. Duda[b], David C. Weindorf[b,*], Bin Li[c], Sourav Roy[a], M.C. Sarathjith[d], Bhabani Sankar Das[a], Laura Paulette[e]

[a] Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, India
[b] Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA
[c] Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA, USA
[d] Precision Farming Development Centre Tavanur, Kelappaji College of Agricultural Engineering and Technology, India
[e] University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania

A R T I C L E I N F O

Handling Editor: A.B. McBratney

Keywords:
Diffuse reflectance spectroscopy
Ordered predictor selection
Variable indicator
Spatial variability
Visible near infrared

A B S T R A C T

Diffuse reflectance spectroscopy (DRS) has shown its potential as a feasible, rapid and non-invasive soil characterization tool. Nevertheless, the use of whole VisNIR spectra in DRS models often incorporates different disruptive and masking effects, eventually producing inefficient model predictions. Thus the careful choice of informative spectral variables is a significant step toward producing robust and useful DRS-based models. This study evaluated the feasibility of combining variable indicator-based DRS outputs and geostatistical interpolations to rapidly produce soil spatial variability maps for six soil properties [sand, clay, silt, total carbon (TC), total nitrogen (TN) and loss-on-ignition organic matter (LOI)]. A total of 300 samples were collected from three catenas of Transylvanian Plain, Romania. First derivative spectra were used to calculate Pearson's correlation coefficient (*r*), biweight midcorrelation (*bicor*), mutual information based adjacency (*AMI*), variable importance in the projection (*VIP*), and their combinations. This variable indicator suite was combined with an ordered predictor selection (OPS) method to choose the optimum number of spectral variables (NSV). This method was tested with partial least squares regression (PLSR) and support vector regression (SVR) with independent validation. Results indicated that the variable indicator-based SVR model yielded superior predictability relative to full-spectrum PLSR model for all soil parameters. Moreover, both PLSR and SVR optimal models used the identical best variable indicators. While *AMI* appeared as the best indicator for four soil attributes (clay, TN, TC and LOI), *bicor* was selected as the best indicator for sand and silt. Spatial variability mapping using optimal SVR model outputs satisfactorily demonstrated management and landscape dynamics across the catenas like the place of manure stockpiles. Summarily, the results of this study indicated that a successful combination of OPS-based variable indicators and their subsequent incorporation into DRS-based chemometric models can potentially improve model predictions that can be further combined with geostatistical interpolation methods to produce spatial maps of soil properties.

1. Introduction

Despite the importance of soil in the agro-ecological environment, describing it for a specific landscape along with a suite of soil attributes is a challenging job owing to its great variability. Notwithstanding the tremendous progress in soil science over the last 20–30 years, the ease of access to soil spatial information varies across the globe. Since the early 1990s, the widespread use of technologies like the global positioning system (GPS), geostatistics, chemometrics, geographic information system (GIS), existing soil maps, satellite and airborne remote sensing, ground spectroradiometers, etc. have jointly evolved into a new approach termed digital soil mapping (McBratney et al., 2003). Although detailed soil maps are available in some countries like Australia and the Netherlands (Bui and Moran, 2001; Hartemink and Sonneveld, 2013), most of the countries in the world do not have detailed soil maps for optimized farm management.

As a result of technological advancements in spectroscopy and chemometrics, visible near-infrared (VisNIR, 350–2500 nm) diffuse

Abbreviations: AMI, mutual information based adjacency; bicor, biweight midcorrelation; DRS, diffuse reflectance spectroscopy; LOI, loss-on-ignition; OPS, ordered predictor selection; PLSR, partial least squares regression; SVR, support vector regression; VIP, variable importance in the projection; VisNIR, visible near infrared

* Corresponding author at: Texas Tech University, Department of Plant and Soil Science, Box 42122, Lubbock, TX 79409, USA.
E-mail address: david.weindorf@ttu.edu (D.C. Weindorf).

https://doi.org/10.1016/j.geoderma.2017.10.043
Received 8 June 2017; Received in revised form 20 October 2017; Accepted 27 October 2017
Available online 21 November 2017

reflectance spectroscopy (DRS) has been widely used in the field of soil science for the last two decades owing to its alacrity, non-invasive nature and ease of analysis (Viscarra Rossel et al., 2006; Duda et al., 2017). This hyperspectral method exhibits the potential to drastically lessen the number of manual steps necessary for traditional soil wet-chemistry analysis, although with a coupled error. The soil VisNIR spectrum is a complex signal and represents a composite response from both molecular vibrations of particular soil components and scattering effects. A host of soil properties have shown good correlation with DRS spectra based multivariate regression model predicted values (Reeves et al., 2000; Reeves et al., 2002; Dunn et al., 2002; Islam et al., 2003; Chakraborty et al., 2010; Chakraborty et al., 2014b; Aldabaa et al., 2015; Chakraborty et al., 2015; Viscarra Rossel et al., 2006). Among different linear and non-linear chemometric algorithms, computationally efficient partial least squares regression (PLSR) has been widely used owing to its capacity to deal with multicollinearity of spectral reflectance values.

As a common practice, use of the whole spectra in DRS models often incorporates surplus information, ultimately producing inefficient model interpretations. Thus scientists have used several variable indicators or 'informative vectors' to select optimum spectral variables to diminish model complexity (Sarathjith et al., 2016). Notably, the efficient use of variable indicators in spectral models has shown to increase model robustness without compromising model accuracy, even increasing it in several occasions (Fernández Pierna et al., 2009). Several studies have proposed a number of sophisticated variable indicators for selection of the optimum number of spectral wavelengths in DRS models like competitive adaptive reweighted sampling (CARS), Pearson's correlation coefficient (*r*), wavelet transformation, genetic algorithms, simulation annealing, moving window PLSR, etc. (Kirkpatrick et al., 1983; Leardi et al., 1992; Ge and Thomasson, 2006; Li et al., 2009; Chen et al., 2011). Further, Teófilo et al. (2009) proposed a fusion of variable indicators with an ordered predictor selection (OPS) method to successfully select the optimal number of spectral variables (NSV). Sarathjith et al. (2014, 2016) proposed two new dependency measures (mutual information based adjacency, *AMI*; and biweight midcorrelation, *bicor*) for the efficient selection of spectral variables. Recently, Vašát et al. (2017a) proposed an ensemble predictive model for more accurate soil organic carbon spectroscopic estimation. Gholizadeh et al. (2013) critically examined the suitability of VisNIR and mid-IR (4000–400 cm^{-1}) spectroscopy as a tool for SOM quantity and quality determination focusing on different pretreatment and calibration procedures and methods. Moreover, Vašát et al. (2014) used peak parameters derived from continuum-removed spectra to predict extractable nutrients in soils with VisNIR DRS. The same authors have used a simple but efficient signal pre-processing called correction by the maximum reflectance for spectroscopic estimation of soil organic carbon (Vašát et al., 2017b).

So far, coarse soil surveys are dependent on conventional soil sampling and subsequent laboratory estimation with negligible emphasis on soil spatial dependence at the *catena* scale. Despite medium to coarse scale area-class soil maps (Scull et al., 2003) which have been utilized in land use and management, their qualitative nature of soil information and coarse scale have caused substantial uncertainty in decision making. Rosemary et al. (2017) revealed a structured variability of topsoil properties in tropical Alfisol *catenas*. However, a thorough understanding of soil properties at the catena scale is critical for farm management as the slopes featured often preclude uniform management practices owing to substantive spatial differences in multiple physicochemical soil properties. Even soil taxonomic classification will vary based on such properties. For example, in the Transylvanian Plain of Romania, Faeozioms and Cernozioms both feature dark soil surface colors, but with subsoil secondary carbonate accumulations at different depths. However, they are intricately interlaced across the landscape, making traditional soil sampling and laboratory characterization impractical. Given that intensive soil sampling and laboratory characterization are very costly, inadequate soil sampling across a large tract of land to characterize soil properties is often associated with substantial errors. As an alternative, several researchers have quantified soil properties and mapped their spatial variability using specific bands from proximal or remote sensing platforms (Schreier, 1977; Aldabaa et al., 2015). For example, Weidong et al. (2002) explored the relationship between soil reflectance in the VisNIR domain (400–2500 nm) and soil moisture using only seven DRS wavebands. Nanni and Demattê (2006) used 22 bands and 13 "Reflectance Inflexion Differences, RID" from different wavelength intervals of the DRS spectrum to successfully predict laboratory soil data ($R^2 > 0.79$). In this study, we intended to evaluate the efficacy of several DRS variable indicators selected via an OPS approach for rapidly mapping soil spatial variability. As such, the objectives of this study were to: a) select the best variable indicators for a host of soil properties for samples collected from three different catenas of Eastern Europe, and b) evaluate the potential of combining variable indicator based chemometric model outputs and geostatistical interpolations to rapidly produce soil spatial variability maps. We hypothesized that geospatial interpolation using DRS variable indicators selected via an OPS approach will produce useful spatial maps of soil parameters.

2. Materials and methods

2.1. General occurrence and soil sampling

For this research, a soil library (300 samples) collected from Romania was used. Three catenas of the Transylvanian Plain (northwest of the country) were evaluated near the towns of Cojocna, Zau de Câmpie, and Căianu, 100 samples were collected from each site (Fig. 1). At each *catena*, sample location was randomly assigned in ArcGIS 10.5 (ESRI, The Redlands, CA, USA), but so as to avoid obvious obstructions such as roads and houses. Then, points were loaded into a handheld GPS for field level geolocation. The area features rolling hills with substantive topographic relief. The Transylvanian Plain features Sarmatian sedimentary rocks, formed of clay in alternation with sands, marls, sandstones, tuffs, and salt and gypsum lenses. Salt deposits vary in thicknesses (up to 1500 m), with surficial expression in Cojocna, Sic, and Turda (Irimuş, 1998). Poorly cemented sands near Zau de Câmpie, Sărmăşel, and Căianu formed domes containing important methane gas resources. Alluvial, deluvial, or colluvial deposits vary in size but have appreciable $CaCO_3$ content and some are saline (Bunescu et al., 2005). More details on climate, topography, and site features can be found in Duda et al. (2017). The collected samples represent a wide variety of soil types and vegetation. Soil types are given both in the Romanian System of Soil Taxonomy (Florea and Munteanu, 2012), as well as in US Soil Taxonomy (Soil Survey Staff, 1999). Some of the most common soil types sampled include preluvosols (Alfisols) and strongly eroded regosols (Entisols) which are common on north-facing slopes. Surface soil (0–5 cm) samples were collected using a standard trowel per Schoeneberger et al. (2012) and geolocated with an eTrex (Garmin Olathe, KS) handheld GPS receiver. Samples were collected in labeled polypropylene bags for transport to the laboratory.

2.2. Laboratory soil analyses

Soils were air-dried and mechanically ground to pass a 2-mm sieve. A modified hydrometer method (Gee and Bauder, 1986) was utilized to determine soil texture with clay determination made at 1440 min with a model 152H soil hydrometer. Soil total carbon (TC) and total nitrogen (TN) values were made using high-temperature combustion via a Truspec® CHN analyzer (LECO Corp., St. Joseph, MI, USA) following the Nelson and Sommers (1996) (Dumas method). Moreover, % soil organic matter was quantified using a muffle furnace as per the loss-on-ignition (LOI) protocol (Nelson and Sommers, 1996).